Integration of single-cell transcriptome and genetic profiles reveals critical cell types and genes in inflammatory bowel disease

Xiaojing Chu[1+*], Siyu Wang[2+], Jiacong Wei[3+], Haosheng Liu[2], Pengcheng Yan[4], Qiuli Yang[1], Yu Tang[5], Bowen Zhang[2], Erli Pang[2*]

[1]The Key Laboratory of Cell Proliferation and Regulation Biology, Ministry of Education, Department of Biology, College of Life Sciences, Beijing Normal University, Beijing 100875, China
[2]MOE Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China.
[3]Department of Pathology, National Cancer Center/ National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021 China.
[4]Chinese institutes for medical research, Beijing 100160, China.
[5]Zhejiang University, School of Medicine, Hangzhou 310058, China.

[+]These authors contributed equally
[*]Correspondence: chuxj@bnu.edu.cn (X.C.), pangerli@bnu.edu.cn (E.P.)

Abstract

The progression of inflammatory bowel disease (IBD), grouped into Crohn's disease (CD) and ulcerative colitis (UC), involves multiple cell types and genes, but the identification of critical cell types and genes that contribute to intestinal inflammation remains challenging. By integrating human intestinal single-cell transcriptomic data from approximately 200 donors, we revealed the stromal cell-derived network in CD, enrichments of immune cells such as Th1 in CD, atypical B cells and Tfh in UC, *AXL*$^+$*SIGLEC6*$^+$ DCs and IgG+ Plasma B cells in both, and the association between *SPP1*$^+$ Macrophages, *FAP*$^+$ Fibroblasts and anti-TNF treatment resistance. We also investigated the gene expression characterization of enteric neurons. Additionally, by linking cell-type–specific transcriptional alterations with genetic profiles of disease risk and intestinal gene expression, we proposed B cells and cycling epithelial cells as critical cell types, and identified epithelial genes, such as *GPX4*, as associated with CD pathology. In summary, our study comprehensively characterized the single-cell landscape of human intestines in IBD and provided insights into cell types and genes highly involved in genetically determined CD pathology, thereby informing CD prediction, diagnosis, and the development of new therapeutic targets.

Introduction

Inflammatory bowel disease (IBD) is the most common intestinal immune-mediated disease grouped into Crohn's disease (CD) and ulcerative colitis (UC). Among them, CD affects both large and small intestines and exhibited higher genetic determinants[1]. UC is limited to the colon, same to colorectal cancer (CRC). Large-scale genome-wide association analyses have identified

hundreds genetic risk loci[2–4], and the enrichments of inflammatory regulators and epithelial function genes highlight the pathological roles of different human intestinal components. Meanwhile, gene expression profiles at the single-cell level have also reported changes in multiple cell-type compositions, covering epithelial, stromal, and immune cells, and uncovered gene programs showing alterations in inflammatory status[5,6] and relevance to clinical treatment[7,8].

However, the previous studies have largely focused on either genetic profiles to map the cell types and genes directly, thus lacking the evaluation of their effective roles in active contexts, or transcriptional alterations, which may not reflect causality in strongly genetically driven diseases. Therefore, it remains a challenge in linking genetic regulation profiles with single-cell transcriptomic profiles to illustrate cellular and molecular mechanisms critical to IBD onset and development in an integrative study.

Integrating large-scale single-cell transcriptomic data greatly enhanced the power in identifying rare cell subsets, and diminished the discrepancy introduced by different analysis procedures from different studies, resulting in a more comprehensive and generalized portrait. Additionally, leveraging the cellular characterization from single-cell transcriptomics, an opportunity was acquired to redefine cellular and molecular involvements for genetic profiles and has achieved remarkable results in limited attempts[9].

In this study, by first integrating single-cell data from multiple datasets, we constructed a harmonized intestinal cell atlas covering samples of different intestinal regions from healthy volunteers, patients with CD and UC. We systematically compared the cellular composition of large and small intestines under different pathological states and characterized the cellular relationships with stromal cells as network hubs that may lead to dysregulation in CD. We revealed the immune cell functional remodeling in broad human intestinal diseases, presenting a full landscape of human intestinal immunity in both hyper-response and suppression. We showed the power of our established atlas in characterizing rare cell types for broad applications. Furthermore, we prioritized cell types and genes critical to CD by integrating cell-type-specific transcriptional alterations with genetic profiles of disease risk and intestinal gene expression. In summary, these findings provide valuable insights into IBD pathology from disease associated cell networks, cell compositions and gene expression alterations, and genetic regulation, which may guide the development of preventive medicine, biomarkers, and new therapeutic targets.

Results

A single-cell atlas of inflammatory bowel disease

To establish a comprehensive single-cell atlas of IBD, we collected published scRNA-seq data covering healthy intestinal tissues, uninflamed and inflamed tissues of CD and UC[5,7,8,10–15], containing 442 samples from either large intestines or small intestines of 196 donors with or without treatment (Fig. 1a,b, Supplementary Table 1). Among the 120 patients, who suffered from CD, 73 patients were reported to be under immune-inhibitor treatment. After quality control largely consistent with our previous study on human colorectum[16], we classified 713,612 cells of high-quality into 71 subsets, including 12 epithelial cell subsets, six endothelial cell

subsets, nine fibroblast or pericyte subsets, two neuronal cell subsets, seven B cell subsets, 16 T cell subsets, four NK/ILC subsets, and 15 myeloid cell subsets, based on their distinct expression patterns (Fig. 1c, Extended Data Fig. 1-3). To address the concerns of batch effects, we showed that the cells were clustered based on cell types rather than datasets (Extended Data Fig. 4a). By further performing a principal component analysis (PCA) on all samples based on their cell subtype proportion in either all non-epithelial cells or all immune cells, the results supported that samples of different datasets, but the same pathological states and sampling regions were clustered together (Extended Data Fig. 4b-d). We also observed that B and T cell subsets contributed the most variation (Supplementary Table 2), in line with their pathological roles in IBD. For the application in a broader field, we established an online browser https://cmb.bnu.edu.cn/Chulab_CD.

Inspired by some cell subsets of different cell lineages clustered together at the transcriptomic level (Fig. 1d), we next investigated the cellular relationships by clustering cell subsets based on their composition in different samples. Consistent with their cellular origins, cell subpopulations from the same lineage tended to be clustered together in all samples (Extended Data Fig. 5a). However, different from the stronger cell occurrence observed in tumor samples reported previously[16], in contrast to the uninflamed samples from CD patients, the cellular occurrence was markedly reduced in inflamed tissues (Fig. 1e). For example, the negative correlations between stromal cells and T cell subsets were largely missing, suggesting a dysregulation in CD.

Further, to infer the potential cause of the dysregulation, we performed a cell communication analysis in each pathological status and more specific interactions were detected in inflamed samples of CD patients (Fig. 2a) than in uninflamed samples of IBD patients, inflamed samples of UC patients as well as samples from healthy donors (Extended Data Fig. 5b), suggesting that a hyper-activated cell network may be associated with the dysregulation in CD. In addition, we identified stromal cells, such as pericytes and fibroblasts, as the major sending cell types with the most ligands expressed, and immune cell types as the major targets in the communications (Fig. 2b, Extended Data Fig. 5c). We also observed that fibroblasts and perivascular cells in the inflamed samples of CD upregulated gene pathways related to stromal cell development and response to TNF signaling (Fig. 2c, Extended Data Fig. 5d). Particularly, LAMB3-mediated cell signaling was identified in the perivascular cells of inflamed samples of CD (Fig. 2d), where *LAMB3* was exclusively expressed (Fig. 2e), in line with the reported LAMB3 effects in promoting intestinal inflammation[17]. We next validated the predicted interactions between *LAMB3* and *CD44* in spatial transcriptomic data of inflamed tissues of CD[6] with a significant colocalization (*Fisher's* exact test, p = 3.20e-13, Fig. 2f and Extended Data Fig. 5e). Together these results revealed a stromal cell-derived cell network in CD, which may lead to the loss of cell occurrence.

Myeloid cells remodeling in intestinal diseases

We next compared the major cell type preference between samples of the large intestine and the small intestine. As shown in Fig. 2g, significant differences were observed in the cell proportion of neurons, T cells and NK/ILCs with larger abundances in small intestine and plasma B cells with larger abundances in the larger intestine. Additionally, the proportion of plasma B cells was

decreased in the inflamed regions compared to controls in large intestine, while increased in small intestine.

Considering the significant influence of sampling regions, we identified cell subsets associated with IBD in only large intestine samples through comparisons conducted among different pathological states. We observed distinct cell composition alterations (Extended Data Fig. 6a, Supplementary Table 3). As presented in Fig. 3a, myeloid cells exhibited a higher proportion in CD but a lower proportion in UC compared to healthy colorectal tissues. In line with the previous result about stromal cell-derived network, we also observed an active cell-cell interactions between myeloid cells and stromal cells (Fig. 3b). Notably, myeloid cells in the inflamed samples of CD patients highly expressed *CXCL8/9* and *CCR1* (Fig. 3c), contributing to an activation of cell communications with stromal cells specific to CD (Fig. 3d). It was reported that CCR1 mediated the accumulation of myeloid cells in tumors[18,19], and the role of CXCL8/9 in enhancing stromal cell proliferation[20]. Together with our observation that myeloid cells in inflamed regions upregulated pathways related to lymphocyte activation (Fig. 3e), we highlighted the pathological roles of myeloid cells in contributing to CD pathology.

By next zooming in on myeloid cell subpopulations, we found that DCs and mast cells drove the observation in all non-epithelial cells, showing enrichments in CD (Fig. 3f, Supplementary Table 4), in line with the previous report[5]. The potential important roles of DCs indicated by the result encouraged us to further investigate the functional preference of myeloid cells in a broad human intestinal disease context. To cope with it, we integrated the IBD atlas with colorectal cancer (CRC) data extracted from our previous established single-cell atlas of human colorectum[16] (Extended Data Fig. 6b). By showing the tissue preference of the myeloid functional subsets well-recognized by our and other previous studies[21,22], we found LAMP3$^+$ DCs were enriched in cancers as well as in CD inflamed tissues (Fig. 3g), supporting the previous reports[23–27]. Studies in tumor microenvironments indicated multiple cell origins of LAMP3$^+$ DCs, corresponding to their diverse function potentials[28]. By aligning LAMP3$^+$ DCs to their two reported major cellular origins, cDC1 and cDC2, we found a higher proportion of cDC1-derived LAMP3$^+$ DCs in CD than in cancers (Fig. 3h), agreeing with the more immune suppressive roles of cDC2-derived LAMP3$^+$ DCs[28]. We also observed more activation of antigen processing functions in cDC1 of CD than in tumors (Extended Data Fig. 6c), which may explain the enrichment of cDC1-derived LAMP3$^+$ DCs in CD.

Additionally, we found that AS DCs, marked by the exclusive expression of *AXL* and *CD5* and the high expression of *CD80*, *SPI1*, *KLF4* and *RUNX3* (Extended Data Fig. 6d), were highly enriched in IBDs than in cancers (Fig. 3g), in line with their reported pro-inflammatory roles[22]. To explore the reason of AS DCs enrichment in inflammation context and consider AS DC was reported as a transiting state between cDC2 and pDC[21], extracellular regulators were predicted by NicheNet analysis[29] on target genes identified using cDC2 and pDCs as controls. We found, for examples, *IFNG*, *IL10* mainly expressed by T/NK and myeloid cells, and *CCL13* mainly expressed by fibroblasts may be responsible for the unique phenotype of AS DCs in CD (Fig. 3i,j and Extended Data Fig. 6e). Among them, fibroblast-*CCL13* has been linked to DCs with pro-inflammation roles in autoimmune diseases[30,31].

Lymphocyte subsets preference in intestinal diseases

Next, we examined the tissue distribution preference of each identified cell subset with Ro/e analysis[32] and observed more universal alterations in the large intestines than that in the small intestines (Fig. 4a, Extended Data Fig. 6f). Notably, different from that in CRC, where stromal cells exhibited large-scale alterations[16], most changes were observed for immune cell components in IBD (Fig. 4a,b), consistent with their dominant contribution to PCA (Supplementary Table 2), indicating the effective roles of immune cells in intestinal inflammation.

It was reported that $CD4^+$ T cells were pathologically involved in both IBD and CRC as key regulators of immune tolerance and inflammation[33–37]. Therefore, we integrated all $CD4^+$ T cells from IBD atlas with $CD4^+$ T cells of paracancerous tissues and tumors from CRC atlas[16] (Extended Data Fig. 7a). As expected, Th1 marked with the expression of *IFNG* was highly enriched in autoimmune (CD), and Treg was enriched in cancer, contributing to an immunosuppressive microenvironment (Fig. 4c). Inspired by the chronic intestinal inflammation may develop into CRC, we hypothesized a transition from other functional $CD4^+$ T cell subsets to Treg, which was supported by the trajectory analysis showing the continuous embedding from Th17 and Tfh to Treg (Fig. 4d). The inferred transiting relationships agree with the reports that Th17 transdifferentiated into Tregs in a mouse autoimmune disease model[38] and Treg transformed into Tfh in a mouse skin inflection model[39].

B-cell components were also largely reprogrammed in both UC and CD inflamed tissues and had crucial functions in intestinal homeostasis[40–42] and tumors[43,44]. Therefore, we redefined the tissue preference in the integrative IBD and CRC data (Extended Data Fig. 7b). We found that atypical B cells (ABCs) and IgG+ Plasma B cells were highly enriched in IBD inflamed regions and tumor tissues in contrast to normal tissues (Fig. 4e). Among them, ABCs upregulated pathways related to B cell activation and antigen signaling (Fig. 4f), agreeing with their proinflammatory roles[45].

The data of anti-TNFa treatment response provided us an opportunity to investigate the cell subset related to immune inhibitor therapy resistance. By comparing with treatment naïve controls, non-responders highly enriched myofibroblast subset mFib-FAP and the macrophage subset Macro-SPP1 (Fig. 4g, Extended Data Fig. 7c). Macro-SPP1 was first identified as a subset of
tumor-associated macrophages[46] with functional potentials in angiogenesis and correlated with worse prognosis and immune-checkpoint blockade treatment resistance[47]. Inflammatory fibroblasts have been linked with anti-TNF resistance[6]. Supporting their potential interactions reported in CRC[48], we observed the colocalization of spatial spots with *FAP* and *SPP1* signatures in CD inflamed tissue (Fig. 4h,i). To sum up, we revealed the remodeling of lymphocyte subsets in human intestinal diseases, may contributing to IBD and anti-TNF treatment resistance.

Characterization of intestinal neurons in inflammation

The crucial roles of neurons in intestinal immune regulations have attracted attention[49], however with few investigations in humans due to the sampling difficulties. Leveraging the single-cell integration technique, we revealed a cluster of intestinal neurons despite largely lost in IBD (Fig.

5a) and may not be recovered after immune inhibitor treatments (Extended Data Fig. 7d). A pathway enrichment analysis on upregulated genes in neurons from inflamed samples implied an enhanced oxidative phosphorylation and immune activation (Fig. 5b), marked by the high expression of class I MHC molecules (Fig. 5c), suggesting their pro-inflammatory roles. Next, we explored the intracellular regulations under these gene expression phenotypes, and uncovered transcription factors such as *STAT1*, *IRF9* that were activated and upregulated in intestinal inflammation (Fig. 5d, Extended Data Fig. 7e).

Considering the role of neurons as a regulation hub, we examined the expression of all glutamate and GABA receptor genes (Fig. 5e). We observed that glutamate AMPA receptor gene *GRIA4* and glutamate kainate receptor gene *GRIK2*, among all other ionotropic glutamate receptor genes, were highly expressed in human intestines, and several GABA receptor genes were also showed remarkable expression levels, in line with their alleviating inflammation in intestines[50]. By revisiting the spatial transcriptomic data of human intestine[6], we validated the expression of *GRIK2*, *GABRA2*, *GABRB1*, *GABRB3*, *GABRE* and *GABBR1* in all spots with neuron signatures (Extended Data Fig. 7f) and showed the colocalization of *GABBR1* with reported top neuron markers in a representative slide as an example (Fig. 5f, Extended Data Fig. 7g). Immunohistochemistry (IHC) staining in human colon tissue samples further validated the expression of two representative neuronal transmitter receptors: GRIA4 and GABBR1 at the protein level (Fig. 5g). Notably, compared to neurons in uninflamed regions, *GRIA4*, *GRIK2* and *GABRB1* were downregulated and *GRIA1*, *GRIN2D*, *GABRB2*, *GABBR1* and *GRM8* were upregulated in inflamed regions (Fig. 5h), implying their involvement in neuron-immune interactions. The detection on gene signatures related to glutamate and GABA synthesis and secretion mirrored altered glutamate and GABA release in inflammation (Fig. 5i, Extended Data Fig. 7h). To sum up, our integrative single-cell atlas offered an opportunity to characterize rare cell type in human intestine.

B cells and epithelial cells as CD critical cell types

Considering CD as a genetics-driven disease, we next aimed to identify cell types crucial to CD. By integrating GWAS profiles and single-cell profiles, we first applied two frameworks, as illustrated in Fig. 6a,b (also see Methods). a: (1) Cell type−specific genes and genes differentially expressed in CD compared to healthy controls (Disease-associated genes) were identified. (2) An enrichment analysis was performed on GWAS profiles as indicated by explained heritability conditional on single nucleotide polymorphism (SNP)s physically located in cell type−specific genes or disease-associated genes. b: (1) We mapped CD GWAS SNPs to genes according to different criteria. (2) Significant overlaps with cell type specific and disease associated genes were detected. Additionally, accentuating disease-associated transcriptional alteration, we also applied an approach, which was adjusted based on our previous study[16] (Methods). Simplified as c: transcriptional alteration levels, defined by distribution deviation in differential expression analysis, were also estimated for all expressed genes and other predefined gene lists.

Consistent with the report on UC[9], B cells were prioritized as the top associated cell types in CD (Fig. 6c) using cell type specific genes in the conditional heritability analysis, however inconsistent results were acquired with disease-associated genes (Extended Data Fig. 8a),

supporting NK/ ILCs as critical cell types associated with the genetic risk of CD. Following the steps of framework b, we first identified epithelial cells and B cells as the top effector cell types with the most significant enrichment of GWAS loci genes (Fig. 6d). Then the distribution analysis in both all expressed genes and GWAS loci genes consistently supported the large alteration in CD epithelial cells (Fig. 6e, Extended Data Fig. 8b). Next, by zooming into epithelial cell subsets, framework b was applied on the combinations of cell type−specific genes and disease-associated genes with loci genes and genes covering either intestinal expression quantitative trait loci (eQTL) or SNPs in enhancer regions identified in human intestinal samples (Fig. 6f). Overlaps with cell type specific genes consistently supported a high relevance of genetic risk genes to cycling epithelial cells. Analysis on disease-associated analysis (Fig. 6f), together with distribution analysis on differential expression analysis profiles with both GWAS loci genes (Fig. 6g) and eQTL genes (Extended Data Fig. 8c) proposed TA, intestinal stem cells and enterocytes exhibited enriched transcriptional alterations in CD. Summarizing from all evidence, we proposed B cells and epithelial cells as critical cell types in CD.

Identification of CD critical genes

Prioritizing CD-associated genes is an important task for both understanding CD pathological mechanisms and finding therapeutic targets. We followed two routes starting from either GWAS genes or cell type and disease genes (Fig. 7a,b). By intersecting GWAS loci genes with both cell type–specific and disease-associated genes, we acquired 185 unique genes for different cell types. Considering the previous clues of CD critical cell types, we focused on B cells and epithelial cells in the further filtering. Leveraging the information of membrane protein, eQTLs and enhancer regions, we proposed five B cell genes and two epithelial genes with top priority in CD pathology (Fig. 7a, Supplementary Table 5). Among them, *ICAM3* has been associated with diffuse large B-Cell lymphoma[51], *CCR10* expressed on plasma cells was involved in CRC[52], colonic homeostatic regulation[53], B cell epithelial cell interactions[54] and virus infection[55]. *GPX4* was reported to regulate mucosal epithelial cell ferroptosis[56,57], supporting their potential inflammatory roles in CD. Lymphotoxin beta receptor (*LTBR*), linked to important intestinal epithelial immune functions[58], has also been prioritized. Further, supporting their associations between genetic profiles of CD and *GPX4*, *LTBR* expression in ileum tissues, we observed strong colocalizations in the regions around rs4806969 (*GPX4*) and rs7954567 (*LTBR*), with H4 = 0.792 and H3+H4 = 0.958, respectively (Fig. 7c). Additionally, considering the high relevance of metabolic pathways to CD and genetic factors regulating metabolites levels enriched for nonsynonymous variants[59], we also provided a gene set covering either variants of nonsynonymous SNV, stop-gain, stop-loss or frameshift substitutions (Supplementary Table 5).

In the second approach, we first applied non-negative matrix factorization (NMF) to identify the core gene programs in CD inflamed tissues and matched each program to their involved cell types. In line with their cellular origins, cell subsets of the same cell lineages were clustered together (Extended Data Fig. 8d). For example, T cells and NK/ILC shared the same program associated with regulation of immune effector process and epithelial subsets enriched for gene programs related to digestion and hormone metabolic processes (Fig. 7d). Subsequently intersecting genes contributing to gene programs with cell type–specific genes and disease-associated genes, we acquired 700 genes with potential functional involvement in CD. Next, we kept 89 genes belonging to the top module pathway and coding membrane or secretory proteins

(Fig. 7b, Supplementary Table 6). Notably, digestion related genes *MUC2*, *MUC4*, *SPINK1*, *TFF3* and *PRSS3*, and metabolism-related genes, *UGT2A3*, *UGT2B17*, *MME*, *AKR1B10*, *DHRS11*, *HSD17B2*, *ACE2* and *ENPEP* popped up as the most important functional genes for intestinal epithelial cells.

Discussion

In this study, we established a high-quality single-cell atlas for human intestine tissues. With the data, we comprehensively investigated the differences between large and small intestines, and more importantly, cell states contributed to different clinical inflammatory phenotypes of IBD. Leveraging the single-cell transcriptional-level information, we illustrated the cellular and molecular mechanisms behind genetic regulation of CD risk and intestinal gene expression and provided rich computational evidence supporting the prioritization of CD critical cell types and genes, which may serve as treatment targets. For example, the pathological role of *GPX4* and *LTBR*, highlighted in our study, have been validated in mouse models. It was reported that an inflammatory response of Gpx4-deficient intestinal epithelial cells could be evoked by ω-6 polyunsaturated fatty acid[60] and wild-type mice given LTBR inhibitors had an increased tumor burden in DSS induced colon cancer[61]. These examples support the value of our prioritization in IBD treatment targets identification.

The single-cell atlas contains >700,000 cells expressing >800 genes of 442 samples from 196 donors. Seventy-one cell subsets were annotated with distinct expression patterns. As a parallel study after CRC integrative atlas[16], applying the same integrative and annotation strategies, our results presented a full blueprint of human intestinal immune-mediated diseases. It is interesting that highly contrasting results were observed for both cell occurrence and cell subset preference in CD and CRC, where the former represents a hyper-response context, while the latter represents an immune suppressive microenvironment. CRC was marked with enhanced cell occurrence in tumors compared to normal tissues[16], whereas we reported reduced cell occurrence in CD. We speculated that cancer cells may force the other local cell types resulting in the observed phenotype by increasing cross cell type interactions. Therefore, it is surprising that a highly interactive yet not-sparse cell network was also found for CD. Together with the results that pericytes and fibroblasts served as the major senders in intercellular communication, we proposed that the acquired additional signaling from stromal cells caused dysregulation in CD.

Different from the stromal cell components that exhibited the largest alteration in CRC[16], a strong cell functional subset remodeling was observed for immune cells in IBD, in line with that the immune cells being identified as the major target cells in the cell communication. We revealed the enrichment of Th1, *AXL*$^{+}$*SIGLEC6*$^{+}$ DCs (AS DCs), cDC1-derived LAMP3^{+} DCs and ABCs in IBD and more importantly, SPP1^{+} macrophages colocalized with FAP^{+} fibroblasts related to anti-TNF treatment resistance. The other explorations upon intra- and intro- cellular signaling provided references related to human intestinal immune regulations.

We took the advantage of integrative analysis in characterizing a rare but important cell type. Neuro-immune interactions have attracted extensive attention but limited by the sampling in human[6,62]. We have not only presented the proinflammatory states of neurons but also uncovered

their expression patterns of major neuronal receptors in the human intestines, providing a window for understanding neuron-immune circuits in tissues.

The investigations of CD critical cell types and genes resulted from a systematic integration of multiple cell type specific transcriptomic and genetic profiles. Despite the fact that these pieces of evidence may point to different conclusions. We believe our efforts largely facilitated the comprehensive understanding of the cellular and molecular mechanisms under CD.

We must acknowledge possible confounders and limitations in this study. First, CD commonly occurs in early childhood, while the incidence of CRC increases along with aging. Thus, age might confound our comparisons between CD and CRC, which may see improvement in future studies. Second, sampling region (epithelial or lamina propria layers) could influence the cell composition of the samples. However, considering the commonly mixing of sampling, we evaluated their general characterization without controlling the sampling layers, which may introduce unwanted bias. Third, different from other cell subsets with exclusive expression patterns, the annotation of Tfr (n = 61 cells) was based on the co-expression of limited Tfh and Treg markers, leading to less confidence. Thus, we did not put any focus on this cell subset in this study and suggest a cautious interpretation for future applications. Lastly, minor effects such as genetic structural variation among different ancestries were not fully controlled, appealing for an in-depth investigation in future studies.

In conclusion, our study comprehensively characterized the single-cell landscape of IBD and proposed critical cell types and genes associated with CD pathology, providing insights in both cellular and molecular mechanisms of IBD and human intestinal immunity. The results may guide the development of preventive medicine, biomarkers, and new therapeutic targets for intestinal immune-mediated diseases.

Methods

Single-cell RNA-seq data collection and quality control

Single-cell transcriptomic data of human intestines in healthy, *Crohn's* disease (CD) and ulcerative colitis (UC) with both inflamed regions and their self-controls were collected from nine studies covering 442 samples from 196 donors[5,7,8,10–15]. The colorectal cancer data with cell type annotation were extracted from our study as described previously[16].

To accurately annotate cell subsets and be consistent to our previous study[16], we first excluded cells with less than 800 detected genes or more than 20% mitochondrial counts. Next, we filtered out potential doublets, which were identified by co-expression of different well-known cell type markers, for unsupervised clustering and cell type identification.

Clustering and cell type identification

Seurat (v5.1.0)[63] and Harmony (v0.1.1)[64] were applied on the expression matrix following the steps described below.

1) Data split by donors (for major cell type identification) or datasets (for subset identification) were normalized and scaled with default settings. We considered top shared 2000 variable genes

for principal component analysis (PCA), and the first 50 principal components (PCs) were corrected with Harmony to regress out the variation introduced by donor-specific effects.
2) The corrected PC 1-50 were used to construct a K nearest neighbor (KNN) graph with default settings and to identify unsupervised cell clusters. Resolutions ranging from 0.1 to 1 were chosen for different cell cluster identification. Cluster-specific gene markers were identified by a "FindMarkers" function in Seurat with adjusted p value < 0.05, min expressing percent > 0.25 and log2 fold change > 0.5.
3) A Uniform Manifold Approximation and Projection (UMAP) was implemented for dimension reduction on 1-50 "Harmony" embedding to visualize cell clustering.
4) Cell types were identified mainly based on the expression pattern of canonical markers for major cell types and well-established or newly identified markers for cell subsets. We took cell annotations taken from each of the collected datasets as a reference for major cell type confirmation.

We annotated cell types in several rounds as described previously[16]. We first carried out this workflow on all single cells to identify major cell types. Then we applied the workflow again on each major cell types to construct cell embeddings for subset identification. In total, we identified 71 subsets in 713,612 cells for downstream analysis.

For integrative analyses on inflammatory bowel disease and colorectal cancer, we first extracted the major cell type of interests based on their original cell type annotations. Then we merged the raw data and applied the same workflow mentioned previously to re-identify the cell subsets.

Cell type abundance evaluation and comparison

Samples with more than 200 cells in total have been considered in this analysis. Cells belonging to each major cell type were divided by the total non-epithelial cells or total immune cells for immune cell subset analysis. A *Student's* t-test (two-tailed) was applied to compare the cell type abundance between samples of different groups.

Tissue preference evaluation of cell subsets

To detect the tissue preference of each subset, we calculated the observed and expected cell ratio (*Ro/e*), where the expected cell numbers were obtained from the chi-square test as previously described[32,65].

Cell origins prediction of LAMP3+ DCs

To infer the fraction of LAMP3+ DCs from different origins, we used SingleR (v2.4.1)[66] to evaluate the similarity between LAMP3+ DCs and cDC1 or cDC2, respectively. The top 50 marker genes of each reference cell type were taken into consideration and cells with max prediction scores > 0.3 were kept for downstream analyses.

Differential expression and pathway enrichment analysis

To perform differentially expression analysis, we applied *Wilcoxon's* rank sum tests on the genes expressed in more than 25% of cells in either group being compared. Genes with adjusted p value < 0.05 and absolute log2 fold change > 0.5 were considered as differentially expressed genes. For enrichment analysis, we used an R package clusterProfiler (v4.2.2)[67] to identify over-represented Gene Ontology (GO) terms, and BH-adjusted p values < 0.05 were considered

statistically significant. Cell type specific and disease associated genes were identified by comparing certain cell type with all other cell types and by comparing inflamed samples with healthy tissues, respectively.

Transcription factor regulon analysis

We used the R package Scenic (v1.3.1)[68] to identify regulatory transcriptional factors (also called regulons) of cell subsets. The regulatory activity (measured in AUC) and expression level of the prioritized transcriptional factors in cell subset of interests were compared between different pathological status with *Wilcoxon's* rank sum tests to get inflammation specific regulons.

Genes signature scoring

Neuronal signature genes were acquired from Gene Ontology (GO). We applied "AddModuleScore" function from Seurat package on the normalized expression values of signature genes to get a score for each cell. *Student's* t-tests were applied to compare scores of different groups.

Developmental trajectory inference

To explore the potential transiting relationships between CD4+ T cell subsets, we applied the Monocle3 (v.1.3.7)[69] with default settings on the harmony corrected embeddings constructed following the workflow described previously.

Hierarchical clustering analysis at transcriptomic and cell proportion level

For transcriptomic level analysis, average expression was calculated on the normalized expression level for each cell subset. For subset clustering based on abundance, a relative proportion was calculated by cell counts dividing all non-epithelial cells beforehand. Then, correlation coefficients between each of two cell subsets were calculated with *Spearman's* correlation analysis. We used an R package pheatmap (v1.0.12) to cluster the cell subsets based on Euclidean distances with "ward.D" approach.

Cell-cell interaction identification

An R package CellChat (v1.6.1)[70] was applied to infer cell-cell interactions between "sender" and "receiver" cell types and p values < 0.05 were considered statistically significant. The sum of ligand-receptor pair probabilities was used to identify major sender and target cell types.

NicheNet analysis

To predict ligands driving transcriptional phenotype of AS DCs, NicheNet (v1.2.1)[29] was applied. Significantly upregulated genes in AS DCs using cDC2 and pDCs as controls were considered as gene set of interest and genes expressed in 5% AS DCs were used as background. Differential expressed genes were calculated by comparing AS DCs and controls (adjusted p value < 0.5 and log2 fold change > 0.25).

Establishment of online browser

An open-access web-based visualization portal using Cirrocumulus (https://github.com/lilab-bcb/cirrocumulus) to facilitate interactive exploration of our single-cell atlas.

Spatial transcriptome

Spatial transcriptome data and spot annotations of intestinal tissues of CD were acquired from a public study[6]. Seurat (v5.1.0)[63] was used for analyses and visualization. *Fisher's* exact test was applied to identify the significant colocalization.

Immunohistochemical staining (IHC)

Formalin-fixed paraffin-embedded (FFPE) colon tissue specimens with ganglion cells of patient diagnosed as CRC and UC were provided by the Cancer Hospital of the Chinese Academy of Medical Sciences with ethical approval. Sections from each sample were cut at 4 μm and stained with three commercially available antibodies for GABBR1 (ha722053, Hangzhou Huaan Biotechnology), GRIA4 (46568, Signalway Anibody) and NeuN (GB11138, Servicebio) according to the manufacturer's instructions.

CD GWAS profiles, SNP annotation

Recent CD GWAS profile with large sample size and most of the reported significant loci[4] was acquired from GWAS-catalog[71] (March 2025) (https://www.ebi.ac.uk/gwas/) under study ID GCST90446794. FUMA (v1.5.2)[72] was applied to map SNPs to genes using 1000G (Phase 3) as reference. SNPs with r2 > 0.6 were considered as in linkage disequilibrium (LD), and with p values < 5e-8 were considered as genome-wide significant. SNPs with minor allele frequency < 0.01 were removed from the analysis. A 10-kb window was used to identify genes physically located within the loci. Human intestine eQTL profiles were acquired from GTEx (v8) and CEDAR project[73]. Annotation of enhancer regions identified in all intestinal samples was acquired from Roadmap[74]. Annovar (2017)[75] was applied to annotate nonsynonymous variants.

Conditional heritability enrichment analysis

SNPs within cell type specific and disease associated genes regions were used for S-LDSC[76] analysis on a CD GWAS profile (GCST90475318) of European ancestry[77]. The enrichment was estimated as explained heritability divided by number of involved SNPs.

Transcriptional alteration estimation

To diminish the effect introduced by different sample size of different cell types and different samples, average expressions at patient level were calculated for each cell type under either inflamed or uninflamed status. Genes expressed in more than 25% patient-status pairs were kept for differential analyses using *Wilcoxon's* rank sum tests and distributions of p values were visualized in a quantile–quantile plot. We estimated the distribution deviation with inflation factor as described in[16].

Genetic colocalization analysis

An R package coloc (v5.2.3)[78] was applied for colocalization analysis. The eQTL summary statistics was acquired from eQTL-Catalog (May 2025) (https://www.ebi.ac.uk/eqtl/), under study ID: QTD000074[73].

NMF in CD inflamed tissues

A consensus NMF package, cNMF (v1.7)[79] has been applied on the averaged expression level of genes in all cell subsets. A tradeoff on the number of components was made between stability and error ending up with 11 components. Top 200 genes with greater contribution for each gene expression program were considered for pathway enrichment analysis.

Other statistical analyses

PCAs analyses on samples of different tissues were performed based on cell subset proportion using "prcomp" function in R. *Fisher's* exact tests were applied to identify significant overlaps between GWAS genes, cell type–specific genes and disease-associated genes.

Visualization

We used the R package ggplot2 (v3.4.2) and Seurat to perform most visualizations, and the package pheatmap was used to generate most heatmaps.

Data availability

Data included in this paper were acquired from NCBI's Gene Expression Omnibus (GSE134809, GSE150115, GSE202052, GSE231993, GSE260842, GSE266616), Single cell portal (https://singlecell.broadinstitute.org/single_cell, SCP1884, SCP259) and upon request from authors. To facilitate the utilization of our IBD atlas for the broader research community, the data has been shared on figshare (Publicly accessible after publication), and an online interactive portal (https://cmb.bnu.edu.cn/Chulab_CD) was developed.

Code availability

This study did not generate any algorithm or software. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Acknowledgments

We are grateful to Dr. Y. Miao and Dr. G. Liu for discussion. The presented study is supported by the Fundamental Research Funds for the Central Universities (Grant 10300310400209510) for X.C.

Author contributions

X.C. designed the study. X.C. and S.W. collected the data and performed bioinformatic analysis supervised by E.P. and with the help from P.Y. J.W. performed IHC experiment. H.L. constructed the online browser with the support from B.Z. X.C. interpreted the data with the help from Q.Y. and Y.T. X.C. wrote the manuscript with input from all authors. The authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Reference

1. McGovern, D. P. B., Kugathasan, S. & Cho, J. H. Genetics of Inflammatory Bowel Diseases. *Gastroenterology* **149**, 1163-1176.e2 (2015).
2. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979–986 (2015).

3. De Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256–261 (2017).
4. Liu, Z. *et al.* Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat Genet* **55**, 796–806 (2023).
5. Kong, L. *et al.* The landscape of immune dysregulation in Crohn's disease revealed through single-cell transcriptomic profiling in the ileum and colon. *Immunity* **56**, 444-458.e5 (2023).
6. Kong, L. *et al.* Single-cell and spatial transcriptomics of stricturing Crohn's disease highlights a fibrosis-associated network. *Nat Genet* **57**, 1742–1753 (2025).
7. Maddipatla, S. C. *et al.* Assessing Cellular and Transcriptional Diversity of Ileal Mucosa Among Treatment-Naïve and Treated Crohn's Disease. *Inflammatory Bowel Diseases* **29**, 274–285 (2023).
8. Martin, J. C. *et al.* Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493-1508.e20 (2019).
9. Jagadeesh, K. A. *et al.* Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat Genet* **54**, 1479–1492 (2022).
10. Li, G. *et al.* Identification of Novel Population-Specific Cell Subsets in Chinese Ulcerative Colitis Patients Using Single-Cell RNA Sequencing. *Cellular and Molecular Gastroenterology and Hepatology* **12**, 99–117 (2021).
11. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).

12. Mitsialis, V. *et al.* Single-Cell Analyses of Colon and Blood Reveal Distinct Immune Cell Signatures of Ulcerative Colitis and Crohn's Disease. *Gastroenterology* **159**, 591-608.e10 (2020).
13. Du, J. *et al.* Selective oxidative protection leads to tissue topological changes orchestrated by macrophage during ulcerative colitis. *Nat Commun* **14**, 3675 (2023).
14. Levantovsky, R. M. *et al.* Multimodal single-cell analyses reveal mechanisms of perianal fistula in diverse patients with Crohn's disease. *Med* **5**, 886-908.e11 (2024).
15. Zhao, Y. *et al.* Multiomic analysis reveals cellular, transcriptomic and epigenetic changes in intestinal pouches of ulcerative colitis patients. *Nat Commun* **16**, 904 (2025).
16. Chu, X. *et al.* Integrative single-cell analysis of human colorectal cancer reveals patient stratification with distinct immune evasion mechanisms. *Nat Cancer* **5**, 1409–1426 (2024).
17. Liu, F. *et al.* LAMB3 Promotes Intestinal Inflammation Through SERPINA3 and Is Directly Transcriptionally Regulated by P65 in Inflammatory Bowel Disease. *Inflammatory Bowel Diseases* **30**, 257–272 (2024).
18. Hirai, H. *et al.* CCR1-mediated accumulation of myeloid cells in the liver microenvironment promoting mouse colon cancer metastasis. *Clin Exp Metastasis* **31**, 977–989 (2014).
19. Zilio, S., Bicciato, S., Weed, D. & Serafini, P. CCR1 and CCR5 mediate cancer-induced myelopoiesis and differentiation of myeloid cells in the tumor. *J Immunother Cancer* **10**, e003131 (2022).
20. Złotkowska, A. & Andronowska, A. Modulatory effect of chemokines on porcine endometrial stromal and endothelial cells. *Domestic Animal Endocrinology* **72**, 106475 (2020).

21. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).
22. Ma, T. *et al.* Pan-cancer Analyses Refine the Single-Cell Portrait of Tumor-Infiltrating Dendritic Cells. *Cancer Research* https://doi.org/10.1158/0008-5472.CAN-24-3595 (2025) doi:10.1158/0008-5472.CAN-24-3595.
23. Maier, B. *et al.* A conserved dendritic-cell regulatory program limits antitumour immunity. *Nature* **580**, 257–262 (2020).
24. Zhang, Q. *et al.* Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* **179**, 829-845.e20 (2019).
25. Yao, R. *et al.* Single-cell transcriptome profiling of the immune space-time landscape reveals dendritic cell regulatory program in polymicrobial sepsis. *Theranostics* **12**, 4606–4628 (2022).
26. Alshevskaya, A. *et al.* Molecular Signatures of Dendritic Cell Activation upon TNF Stimulation: A Multi-Omics Study in Rheumatoid Arthritis. *IJMS* **26**, 6071 (2025).
27. He, H. *et al.* Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis. *Journal of Allergy and Clinical Immunology* **145**, 1615–1628 (2020).
28. Cheng, S. *et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792-809.e23 (2021).
29. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* **17**, 159–162 (2020).
30. Ma, F. *et al.* Single cell and spatial sequencing define processes by which keratinocytes and fibroblasts amplify inflammatory responses in psoriasis. *Nat Commun* **14**, 3455 (2023).

31. Mitamura, Y. *et al.* Spatial transcriptomics combined with single□cell RNA □sequencing unravels the complex inflammatory cell network in atopic dermatitis. *Allergy* **78**, 2215–2231 (2023).
32. Guo, X. *et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* **24**, 978–985 (2018).
33. Kosinsky, R. L. *et al.* The FOXP3+ Pro-Inflammatory T Cell: A Potential Therapeutic Target in Crohn's Disease. *Gastroenterology* **166**, 631-644.e17 (2024).
34. Yokoi, T. *et al.* Identification of a unique subset of tissue-resident memory $CD4^+$ T cells in Crohn's disease. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2204269120 (2023).
35. Zundler, S. *et al.* Hobit- and Blimp-1-driven CD4+ tissue-resident memory T cells control chronic intestinal inflammation. *Nat Immunol* **20**, 288–300 (2019).
36. Thoma, O.-M. *et al.* p21 Prevents the Exhaustion of CD4+ T Cells Within the Antitumor Immune Response Against Colorectal Cancer. *Gastroenterology* **166**, 284-297.e11 (2024).
37. Li, M. *et al.* Identification and validation of a CD4+ T cell-related prognostic model to predict immune responses in stage III-IV colorectal cancer. *BMC Gastroenterol* **25**, 153 (2025).
38. Gagliani, N. *et al.* Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature* **523**, 221–225 (2015).
39. Gribonika, I. *et al.* Skin autonomous antibody production regulates host–microbiota interactions. *Nature* **638**, 1043–1053 (2025).
40. Spencer, J. & Sollid, L. M. The human intestinal B-cell response. *Mucosal Immunology* **9**, 1113–1124 (2016).

41. Uzzan, M. *et al.* Ulcerative colitis is characterized by a plasmablast-skewed humoral response associated with disease activity. *Nat Med* **28**, 766–779 (2022).
42. Bamias, G., Kitsou, K. & Rivera-Nieves, J. The Underappreciated Role of Secretory IgA in IBD. *Inflammatory Bowel Diseases* **29**, 1327–1341 (2023).
43. Wang, Z. *et al.* Leucine-tRNA-synthetase-2-expressing B cells contribute to colorectal cancer immunoevasion. *Immunity* **55**, 1067-1081.e8 (2022).
44. Ji, L. *et al.* scRNA-seq of colorectal cancer shows regional immune atlas with the function of CD20+ B cells. *Cancer Letters* **584**, 216664 (2024).
45. Holla, P. *et al.* Shared transcriptional profiles of atypical B cells suggest common drivers of expansion and function in malaria, HIV, and autoimmunity. *Sci. Adv.* **7**, eabg8384 (2021).
46. Zhang, L. *et al.* Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell* **181**, 442-459.e29 (2020).
47. Yang, Z. *et al.* Single-cell sequencing reveals immune features of treatment response to neoadjuvant immunochemotherapy in esophageal squamous cell carcinoma. *Nat Commun* **15**, 9097 (2024).
48. Qi, J. *et al.* Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer. *Nat Commun* **13**, 1742 (2022).
49. Wallrapp, A. & Chiu, I. M. Neuroimmune Interactions in the Intestine. *Annual Review of Immunology* **42**, 489–519 (2024).
50. Liu, N. *et al.* Enteric GABAergic neuron-derived γ-aminobutyric acid initiates expression of Igfbp7 to sustain ILC3 homeostasis. *Nat Immunol* **26**, 404–415 (2025).
51. Li, L. *et al.* Functional evaluation of various ICAM3 transcript variants in diffuse large B-Cell lymphoma. *Leukemia & Lymphoma* **63**, 2869–2878 (2022).

52. Xia, J. *et al.* Single□cell landscape and clinical outcomes of infiltrating B cells in colorectal cancer. *Immunology* **168**, 135–151 (2023).
53. Zhao, L. *et al.* Coordinated co-migration of CCR10+ antibody-producing B cells with helper T cells for colonic homeostatic regulation. *Mucosal Immunology* **14**, 420–430 (2021).
54. Xiong, N., Fu, Y., Hu, S., Xia, M. & Yang, J. CCR10 and its ligands in regulation of epithelial immunity and diseases. *Protein Cell* **3**, 571–580 (2012).
55. Nakayama, T. *et al.* Human B Cells Immortalized with Epstein-Barr Virus Upregulate CCR6 and CCR10 and Downregulate CXCR4 and CXCR5. *J Virol* **76**, 3072–3077 (2002).
56. Zhang, J. *et al.* Scutellaria baicalensis Extracts Restrict Intestinal Epithelial Cell Ferroptosis by Regulating Lipid Peroxidation and GPX4/ACSL4 in Colitis. *Phytomedicine* **141**, 156708 (2025).
57. Yoshida, M. *et al.* Involvement of cigarette smoke-induced epithelial cell ferroptosis in COPD pathogenesis. *Nat Commun* **10**, 3145 (2019).
58. Chen, Q. *et al.* LTβR-RelB signaling in intestinal epithelial cells protects from chemotherapy-induced mucosal damage. *Front. Immunol.* **15**, 1388496 (2024).
59. Chu, X. *et al.* Integration of metabolomics, genomics, and immune phenotypes reveals the causal roles of metabolites in disease. *Genome Biol* **22**, 198 (2021).
60. Mayr, L. *et al.* Dietary lipids fuel GPX4-restricted enteritis resembling Crohn's disease. *Nat Commun* **11**, 1775 (2020).
61. Kempski, J. *et al.* IL22BP Mediates the Antitumor Effects of Lymphotoxin Against Colorectal Tumors in Mice and Humans. *Gastroenterology* **159**, 1417-1430.e3 (2020).
62. Drokhlyansky, E. *et al.* The Human and Mouse Enteric Nervous System at Single-Cell Resolution. *Cell* **182**, 1606-1622.e23 (2020).

63. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

64. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).

65. Zhang, L. *et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).

66. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163–172 (2019).

67. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).

68. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).

69. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979–982 (2017).

70. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).

71. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* **47**, D1005–D1012 (2019).

72. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).

73. Momozawa, Y. *et al.* IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat Commun* **9**, 2427 (2018).

74. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

75. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**, e164–e164 (2010).

76. ReproGen Consortium *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228–1235 (2015).

77. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).

78. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet* **10**, e1004383 (2014).

79. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* **8**, e43803 (2019).

Figure Captions

Figure 1

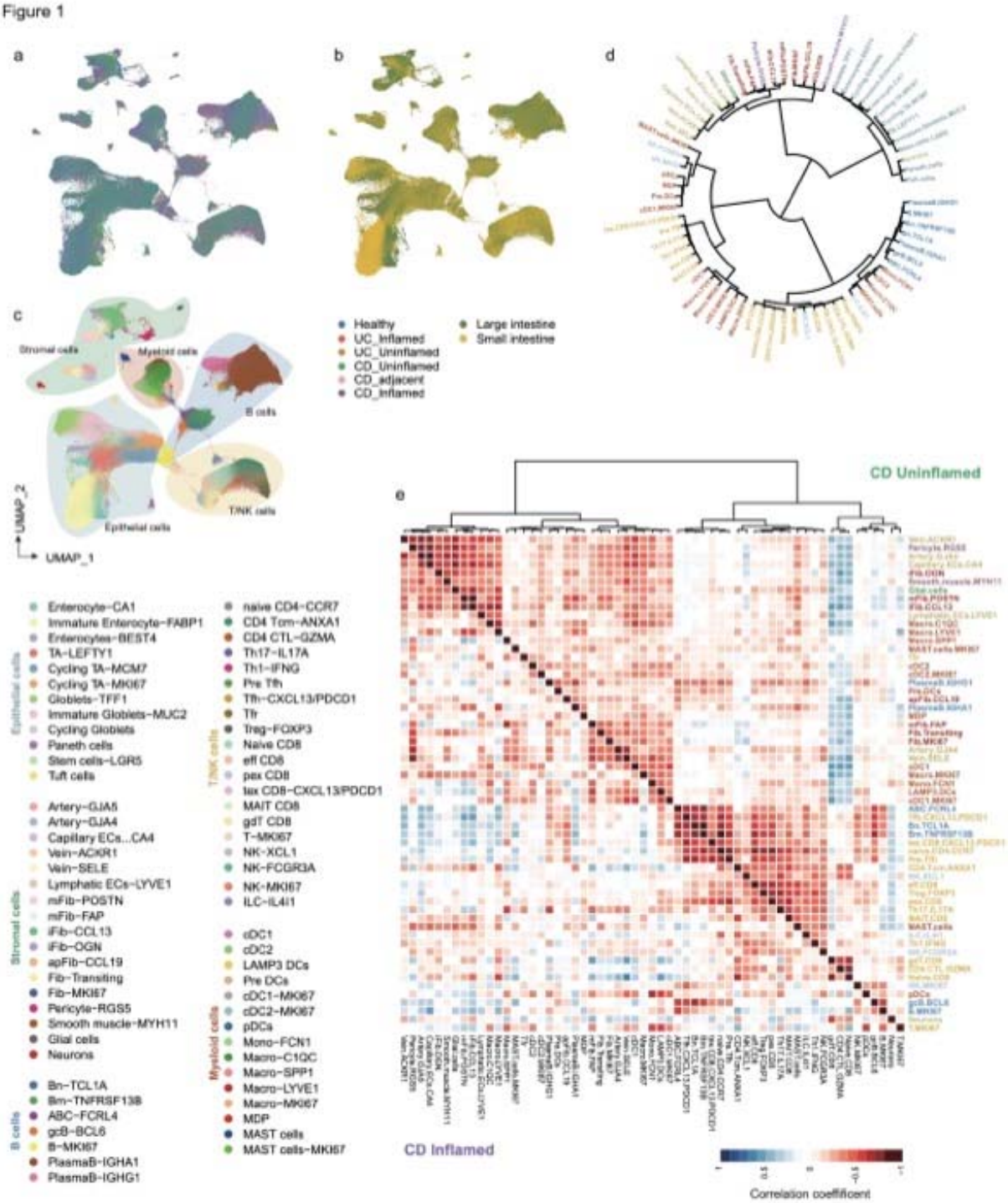

a
b
d
c
Stromal cells
Myeloid cells
B cells
T/NK cells
Epithelial cells
UMAP_2
UMAP_1
Healthy
UC_Inflamed
UC_Uninflamed
CD_Uninflamed
CD_adjacent
CD_Inflamed
Large intestine
Small intestine
e
CD Uninflamed
CD Inflamed
Correlation coefficient
Epithelial cells
Enterocyte-CA1
Immature Enterocyte-FABP1
Enterocytes-BEST4
TA-LEFTY1
Cycling TA-MCM7
Cycling TA-MKI67
Globlets-TFF1
Immature Globlets-MUC2
Cycling Globlets
Paneth cells
Stem cells-LGR5
Tuft cells
Stromal cells
Artery-GJA5
Artery-GJA4
Capillary ECs...CA4
Vein-ACKR1
Vein-SELE
Lymphatic ECs-LYVE1
mFib-POSTN
mFib-FAP
iFib-CCL13
iFib-OGN
apFib-CCL19
Fib-Transiting
Fib-MKI67
Pericyte-RGS5
Smooth muscle-MYH11
Glial cells
Neurons
B cells
Bn-TCL1A
Bm-TNFRSF13B
ABC-FCRL4
gcB-BCL6
B-MKI67
PlasmaB-IGHA1
PlasmaB-IGHG1
T/NK cells
naive CD4-CCR7
CD4 Tcm-ANXA1
CD4 CTL-GZMA
Th17-IL17A
Th1-IFNG
Pre Tfh
Tfh-CXCL13/PDCD1
Tfr
Treg-FOXP3
Naive CD8
eff CD8
pex CD8
tex CD8-CXCL13/PDCD1
MAIT CD8
gdT CD8
T-MKI67
NK-XCL1
NK-FCGR3A
NK-MKI67
ILC-IL4I1
Myeloid cells
cDC1
cDC2
LAMP3 DCs
Pre DCs
cDC1-MKI67
cDC2-MKI67
pDCs
Mono-FCN1
Macro-C1QC
Macro-SPP1
Macro-LYVE1
Macro-MKI67
MDP
MAST cells
MAST cells-MKI67

Fig. 1. Single cell landscape of human intestines. a-c, UMAP plot showing cell clustering colored by tissue pathological states (a), sampling regions (b) and identified subsets (c) (n=713612 cells). fib: fibroblasts, n: naïve, m: memory, cm: central memory, gc: germinal center, macro: macrophage, eff: effector, pex: progenitor exhausted, tex: terminal exhausted. **d**, Hierarchical clustering of cell subsets based on average expression levels in all samples (n = 314 samples). **e**, Hierarchical clustering of cell subsets based on *Spearman's* correlation coefficients between each of two subsets in either CD uninflamed samples (top right, n = 124 samples) or CD inflamed samples (bottom left, n = 68 samples).

Figure 2

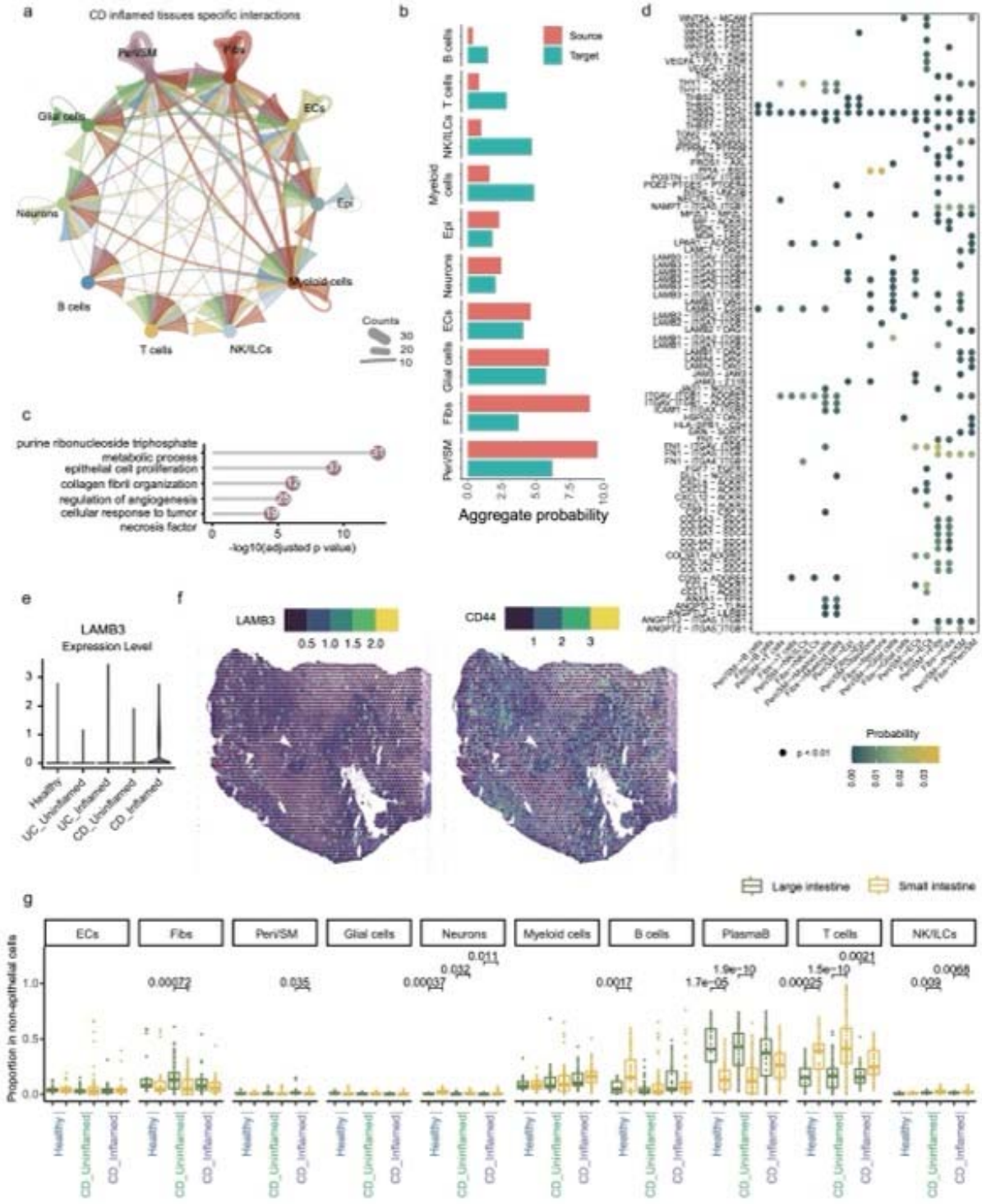
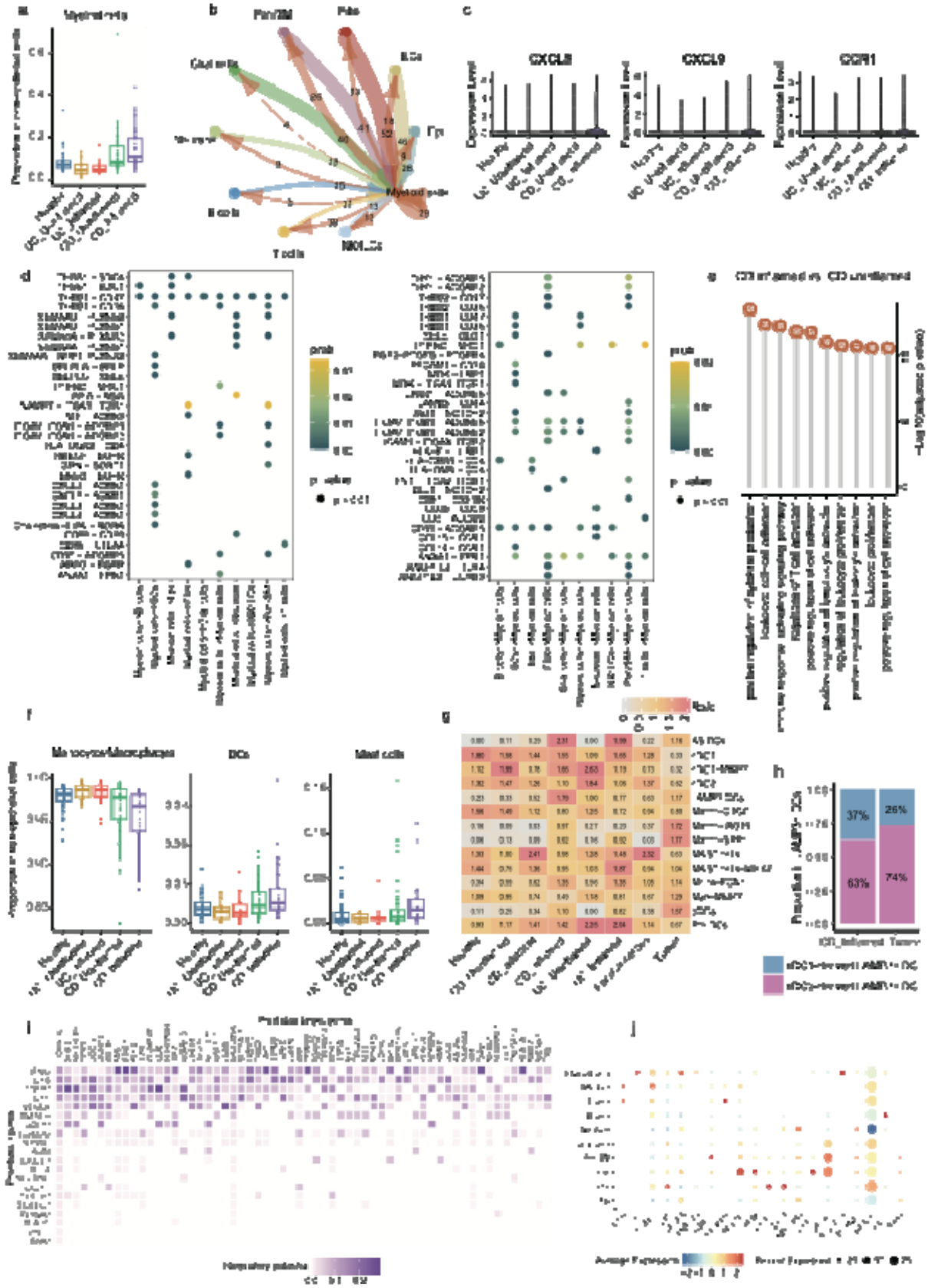

a
CD inflamed tissues specific interactions
Peri/SM
Fibs
ECs
Glial cells
Epi
Neurons
Myeloid cells
B cells
T cells
NK/ILCs
Counts
30
20
10
b
Source
Target
B cells
T cells
NK/ILCs
Myeloid cells
Epi
Neurons
ECs
Glial cells
Fibs
Peri/SM
Aggregate probability
c
purine ribonucleoside triphosphate metabolic process
epithelial cell proliferation
collagen fibril organization
regulation of angiogenesis
cellular response to tumor necrosis factor
-log10(adjusted p value)
d
Probability
e
LAMB3
Expression Level
Healthy
UC_Uninflamed
UC_Inflamed
CD_Uninflamed
CD_Inflamed
f
LAMB3
CD44
g
Large intestine
Small intestine
ECs
Fibs
Peri/SM
Glial cells
Neurons
Myeloid cells
B cells
PlasmaB
T cells
NK/ILCs
Proportion in non-epithelial cells

Fig. 2. Inferred cell interactions in CD inflamed tissues. a, Circle plot showing counts of predicted cell interactions specific to CD inflamed samples. **b**, Barplot showing aggregate probabilities of each cell type as source and target in cell interactions. **c**, Lollipop plot showing upregulated pathways in the inflamed tissues compared to uninflamed tissues in pericytes. Numbers in the dots indicate gene counts matched to corresponding biological pathways. Over-representation analysis (BH adjustment). **d**, Dotplot showing estimated probability of ligand-receptor pairs from fibroblasts and pericytes to other cell types identified in only CD inflamed tissues. **e**, Violin plot showing the expression level of *LAMB3* in pericytes of different tissues (n = 658 (Healthy), 138 (UC_Uninflamed), 278 (UC_Inflamed), 1412 (CD_Uninflamed) and 2036 (CD_Inflamed) cells). **f**, Colocalization of *LAMB3* and *CD44* in inflamed regions revealed by spatial transcriptomics. **g**, Boxplots showing cell type proportion in all non-epithelial cells (n = 21, 26, 40, 84, 27 and 41 from left to right). Center line indicates the median value, lower and upper hinges represent the first and third quartiles, respectively, and whiskers denote 1.5× interquartile range. *Student's* t-test (two-sided, unpaired).

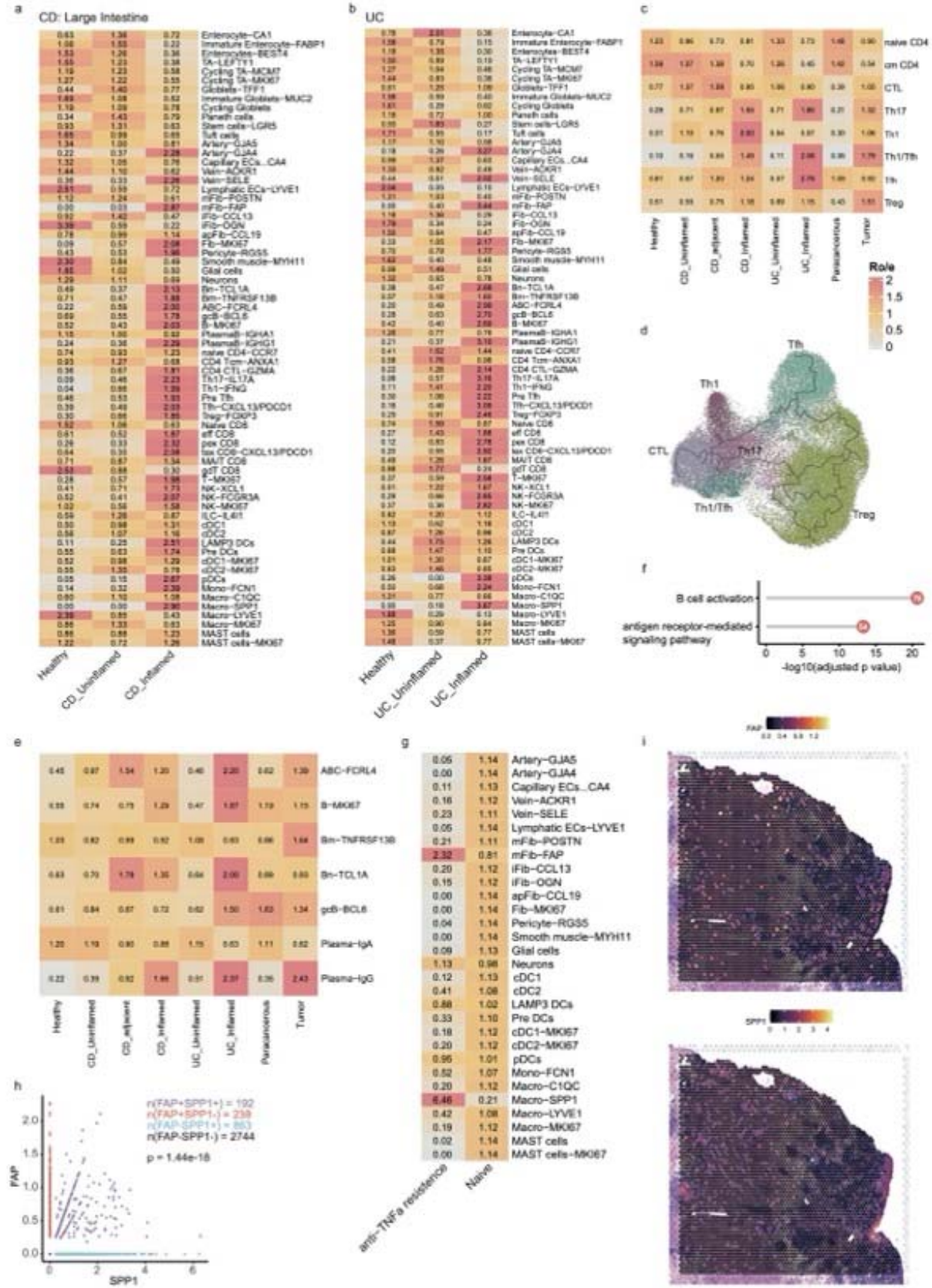
Figure 3


a
b
c
CXCL8
CXCL9
CCR1
d
e
f
Macrophage/Monocytes
DCs
Mast cells
g
h
i
j

790

Fig. 3. Myeloid cells characterization in human intestinal diseases. a, Boxplot showing large intestinal myeloid cell proportion in all non-epithelial cells (n = 45, 14, 18, 40 and 27 samples from left to right). Center line indicates the median value, lower and upper hinges represent the first and third quartiles, respectively, and whiskers denote 1.5× interquartile range. *Student's* t-test (two-sided, unpaired). p values were presented in Supplementary Table 3. **b**, Circle plot showing counts of all predicted interactions around myeloid cells. **c**, Violin plots showing the expression level of *CXCL8/9* and CCR1 in myeloid cells from different tissues (n = 6297, 797, 882, 19353 and 18620 cells from left to right). **d**, Dotplots showing estimated probability of ligand-receptor pairs from (Left) or to (Right) myeloid cells identified in only CD inflamed tissues. **e**, Lollipop plot showing upregulated pathways in the inflamed tissues compared to uninflamed tissues in myeloid cells. Numbers in the dots indicate gene counts matched to corresponding biological pathways. Over-representation analysis (BH adjustment). **f**, Boxplots showing myeloid cell subpopulations proportion in all non-epithelial cells in large intestine samples (n = 45, 14, 18, 40 and 27 from left to right). Center line indicates the median value, lower and upper hinges represent the first and third quartiles, respectively, and whiskers denote 1.5× interquartile range. *Student's* t-test (two-sided, unpaired). p values were presented in Supplementary Table 4. **g**, Heatmap showing tissue preference of each myeloid cell subset estimated by *Ro/e* score. **h**, Stacked bar plots showing the predicted origins of $LAMP3^{+}$ DC in different tissue types. **i**, Heatmap showing the regulatory potentials of the top 20 prioritized ligands of AS DCs. **j**, Dotplot showing the expression levels of the top 20 prioritized ligands in each cell type.

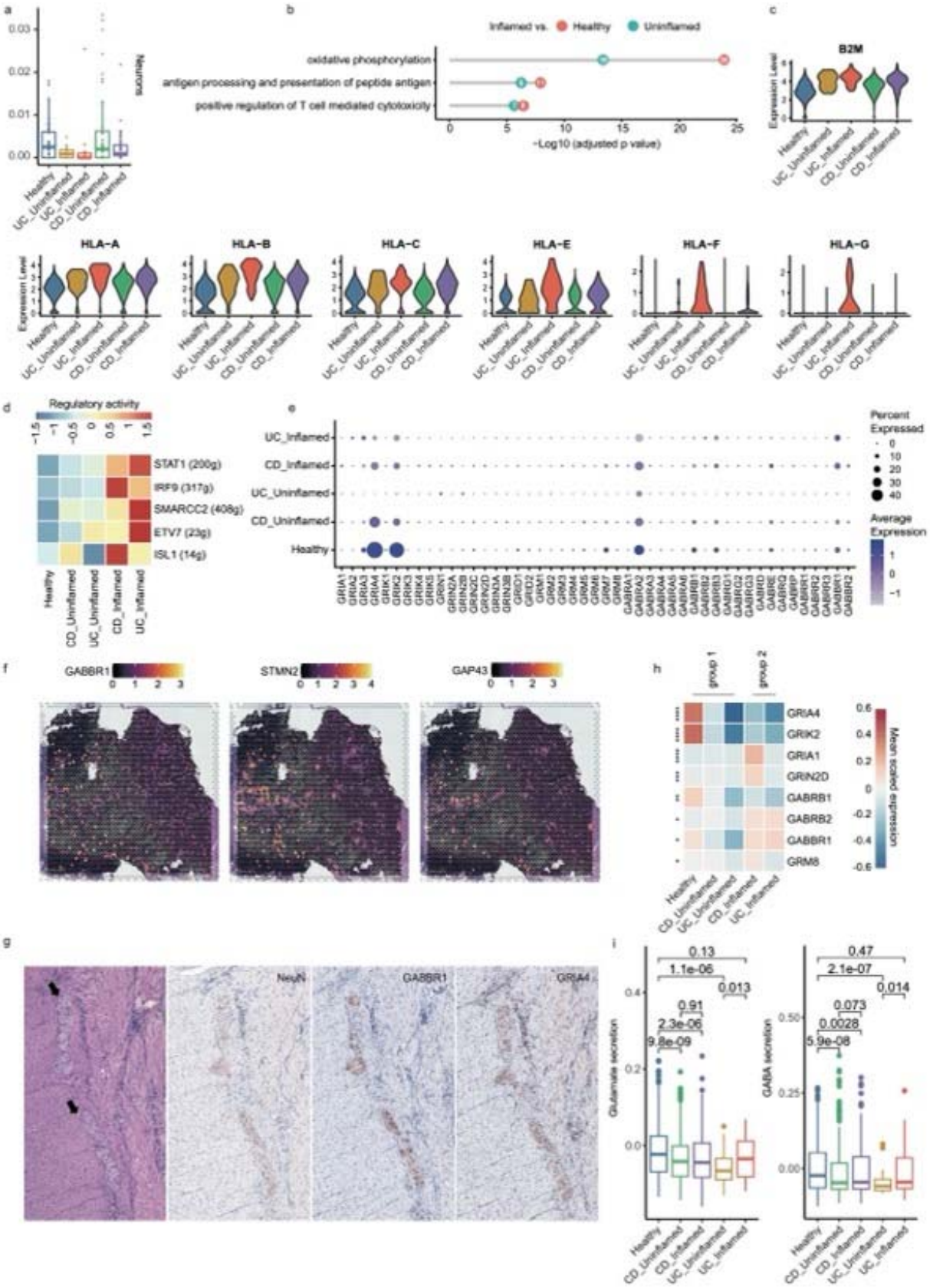
Figure 4


a
CD: Large Intestine
Healthy
CD_Uninflamed
CD_Inflamed
b
UC
Healthy
UC_Uninflamed
UC_Inflamed
c
Healthy
CD_Uninflamed
CD_adjacent
CD_Inflamed
UC_Uninflamed
UC_Inflamed
Paracancerous
Tumor
Role
d
Th
Th1
CTL
Th17
Th1/Tfh
Treg
f
B cell activation
antigen receptor-mediated signaling pathway
-log10(adjusted p value)
e
ABC-FCRL4
B-MKI67
Bm-TNFRSF13B
Bn-TCL1A
gcB-BCL6
Plasma-IgA
Plasma-IgG
g
anti-TNFa resistence
Naive
i
FAP
SPP1
h
n(FAP+SPP1+) = 192
n(FAP+SPP1-) = 239
n(FAP-SPP1+) = 853
n(FAP-SPP1-) = 2744
p = 1.44e-18
FAP
SPP1

813

Fig. 4. Lymphocyte subsets preference in human intestinal diseases. a-c, Heatmap showing tissue preference of each lymphocyte subset estimated by *Ro/e* score. **d**, Predicted cell trajectory of activated $CD4^+$ T cells (n = 67532 cells). **e**, Heatmap showing tissue preference of each B cell subset estimated by *Ro/e* score. **f**, Lollipop plot showing upregulated pathways in ABCs compared to other B cell subsets. Numbers in the dots indicate gene counts matched to corresponding biological pathways. Over-representation analysis (BH adjustment). **g**, Heatmap showing tissue preference of each cell subset estimated by *Ro/e* score. **h**, Scatter plot showing expression of *FAP* and *SPP1* in spatial spots of all slides. *Fisher's* exact test. **i**, Colocalization of *FAP* and *SPP1* in inflamed regions revealed by spatial transcriptomics.

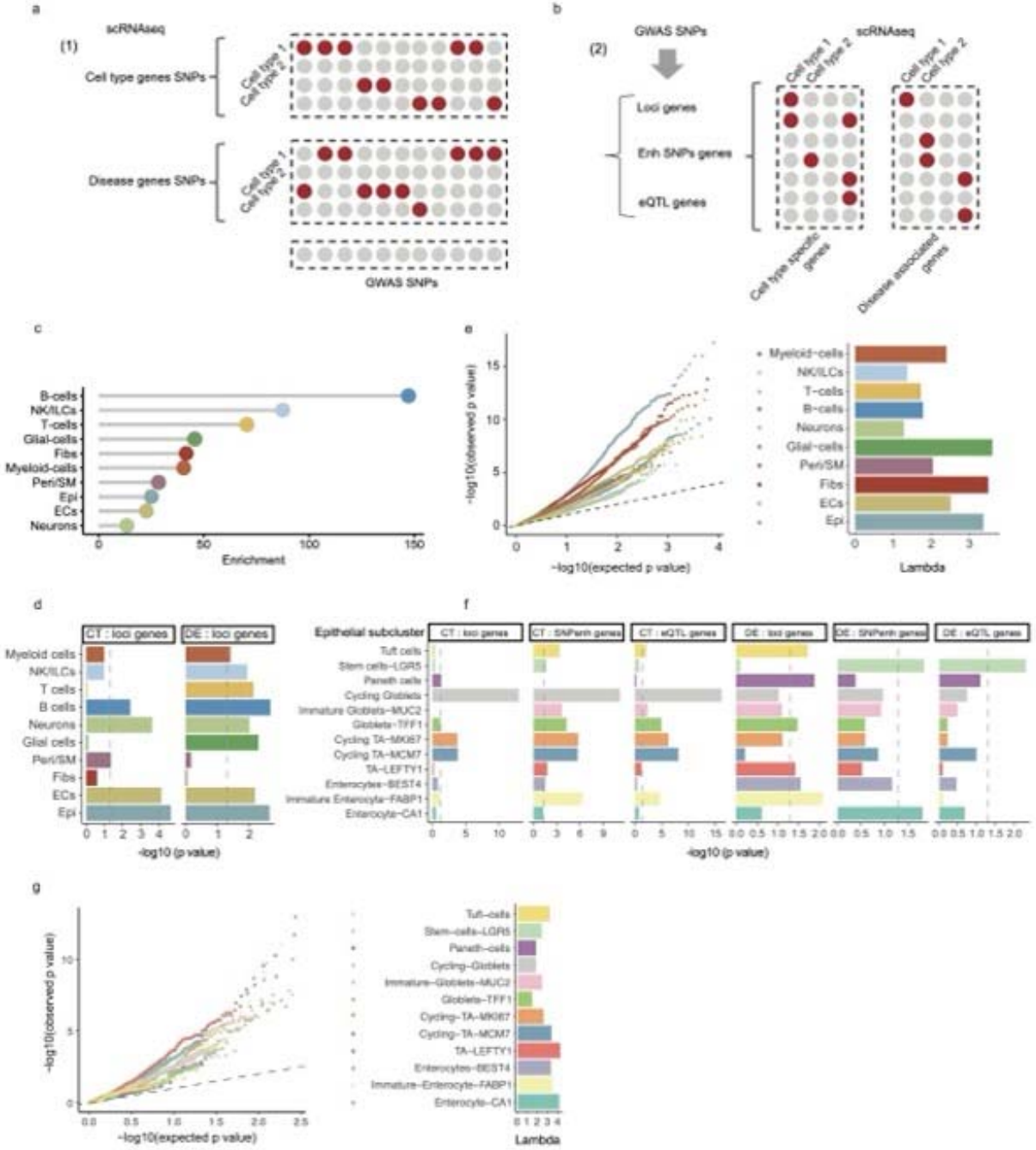
Figure 5


a
Neurons
Healthy
UC_Uninflamed
UC_Inflamed
CD_Uninflamed
CD_Inflamed
b
Inflamed vs.
Healthy
Uninflamed
oxidative phosphorylation
antigen processing and presentation of peptide antigen
positive regulation of T cell mediated cytotoxicity
-Log10 (adjusted p value)
c
B2M
Expression Level
HLA-A
HLA-B
HLA-C
HLA-E
HLA-F
HLA-G
d
Regulatory activity
STAT1 (200g)
IRF9 (317g)
SMARCC2 (408g)
ETV7 (23g)
ISL1 (14g)
e
Percent Expressed
Average Expression
f
GABBR1
STMN2
GAP43
h
group 1
group 2
GRIA4
GRIK2
GRIA1
GRIN2D
GABRB1
GABRB2
GABBR1
GRM8
Mean scaled expression
g
NeuN
GABBR1
GRIA4
i
Glutamate secretion
GABA secretion
0.13
1.1e-06
0.013
0.91
2.3e-06
9.8e-09
0.47
2.1e-07
0.014
0.073
0.0028
5.9e-08

824

Fig. 5. Intestinal neurons characterization. a, Boxplot showing large intestinal neurons proportion in all non-epithelial cells (n = 45, 14, 18, 40 and 27 samples from left to right). Center line indicates the median value, lower and upper hinges represent the first and third quartiles, respectively, and whiskers denote 1.5× interquartile range. *Student's* t-test (two-sided, unpaired). p values were presented in Supplementary Table 3. **b**, Lollipop plot showing upregulated pathways in neurons of inflamed tissues compared to that of uninflamed and healthy tissues. Numbers in the dots indicate gene counts matched to corresponding biological pathways. Over-representation analysis (BH adjustment). **c**, Violin plots showing the expression level of MHC-I genes in neurons of different tissues (n = 714, 32, 37, 986 and 527 cells from left to right). **d**, Heatmap showing regulatory activity of representative transcription factors upregulated in inflamed samples. **e**, Dotplot showing expression levels of neuronal receptors in different tissues. **f**, Colocalization of *GABBR1* and reported top intestinal neuron markers revealed by spatial transcriptomics. **g,** Representative images (200x) of colorectal ganglion cells (indicated by arrows) stained by hematoxylin-eosin (H&E) and IHC. **h**, Heatmap showing expression levels of neuronal receptors differentially expressed in inflamed tissues compared to normal tissues. p values from top to bottom equal to 2.61e-13 (****), 2.67e-9 (****), 4.48e-7 (****), 9.09e-4 (***), 9.13e-3 (**), 3.07e-2 (*), 3.16e-2 (*) and 4.09e-2 (*). **i**, Boxplots showing signature scores of neurons from different tissues (n = 714, 32, 37, 986 and 527 cells from left to right). Center line indicates the median value, lower and upper hinges represent the first and third quartiles, respectively, and whiskers denote 1.5× interquartile range. *Student's* t-test (two-sided, unpaired).

Figure 6


a
scRNAseq
(1)
Cell type genes SNPs
Disease genes SNPs
Cell type 1
Cell type 2
GWAS SNPs
b
GWAS SNPs
(2)
scRNAseq
Loci genes
Enh SNPs genes
eQTL genes
Cell type specific genes
Disease associated genes
c
B-cells
NK/ILCs
T-cells
Glial-cells
Fibs
Myeloid-cells
Peri/SM
Epi
ECs
Neurons
Enrichment
e
-log10(observed p value)
-log10(expected p value)
Lambda
d
CT : loci genes
DE : loci genes
Myeloid cells
NK/ILCs
T cells
B cells
Neurons
Glial cells
Peri/SM
Fibs
ECs
Epi
-log10 (p value)
f
Epithelial subcluster
CT : loci genes
CT : SNPenh genes
CT : eQTL genes
DE : loci genes
DE : SNPenh genes
DE : eQTL genes
Tuft cells
Stem cells-LGR5
Paneth cells
Cycling Globlets
Immature Globlets-MUC2
Globlets-TFF1
Cycling TA-MKI67
Cycling TA-MCM7
TA-LEFTY1
Enterocytes-BEST4
Immature Enterocyte-FABP1
Enterocyte-CA1
-log10 (p value)
g
-log10(observed p value)
-log10(expected p value)
Lambda

847

Fig. 6. Identification of CD critical cell types. a-b, Schematics of CD critical cell type identification frameworks. **c**, Lollipop plot showing enrichment score in sLDSC analysis. **d**, Barplots showing enrichment significances of loci genes in cell type specific (CT) and disease associated (DE) genes. *Fisher's* exact tests. **e**, Left: qqplot showing p value distribution of all expressed genes in the comparisons between CD inflamed tissues and healthy tissues (n = 8535, 2132, 7046, 10055, 2507, 1170, 1679, 8965, 7881 and 11838 genes from top to bottom). Right: Barplot showing lambda values of different cell types. **f**, Barplots showing enrichment significances of loci genes, genes with SNPs located in enhancer regions and eQTL genes in cell type specific (CT) and disease associated (DE) genes. *Fisher's* exact tests. **g**, Left: qqplot showing p value distribution of expressed loci genes in the comparisons between CD inflamed tissues and healthy tissues (n = 45, 342, 31, 183, 375, 315, 380, 406, 396, 195, 281 and 278 genes from top to bottom). Right: Barplot showing lambda values of different epithelial subtypes.

Figure 7

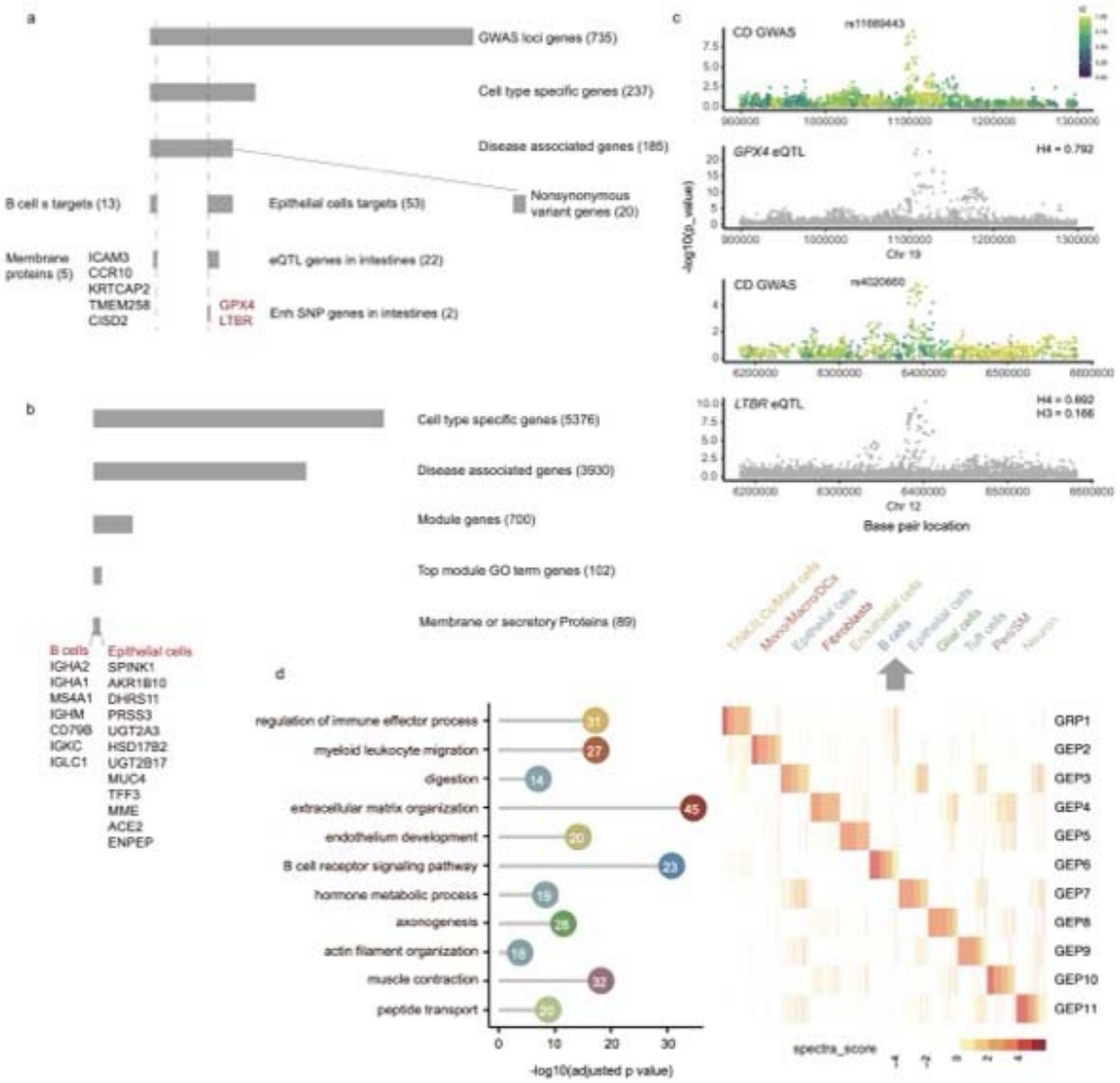

a
GWAS loci genes (735)
Cell type specific genes (237)
Disease associated genes (186)
B cell s targets (13)
Epithelial cells targets (53)
Nonsynonymous variant genes (20)
Membrane proteins (5)
ICAM3
CCR10
KRTCAP2
TMEM258
CISD2
eQTL genes in intestines (22)
GPX4
LTBR
Enh SNP genes in intestines (2)
b
Cell type specific genes (5376)
Disease associated genes (3930)
Module genes (700)
Top module GO term genes (102)
Membrane or secretory Proteins (89)
B cells
IGHA2
IGHA1
MS4A1
IGHM
CD79B
IGKC
IGLC1
Epithelial cells
SPINK1
AKR1B10
DHRS11
PRSS3
UGT2A3
HSD17B2
UGT2B17
MUC4
TFF3
MME
ACE2
ENPEP
c
CD GWAS
rs11689443
GPX4 eQTL
H4 = 0.792
-log10(p_value)
Chr 19
CD GWAS
rs4320660
LTBR eQTL
H4 = 0.892
H3 = 0.166
Chr 12
Base pair location
d
regulation of immune effector process
myeloid leukocyte migration
digestion
extracellular matrix organization
endothelium development
B cell receptor signaling pathway
hormone metabolic process
axonogenesis
actin filament organization
muscle contraction
peptide transport
-log10(adjusted p value)
Fibroblasts
B cells
GRP1
GEP2
GEP3
GEP4
GEP5
GEP6
GEP7
GEP8
GEP9
GEP10
GEP11
spectra_score

Fig. 7. Identification of CD critical genes. a-b, Schematics of CD critical genes identification frameworks. **c**, Scatter plots showing genetic colocalizations between CD GWAS profiles and *GPX4*/ *LTBR* eQTL profiles. **d**, Left: Lollipop plot showing top enriched pathway in each gene expression program (GEP). Right: Heatmap showing spectra scores of top 200 GEP genes.